2018-08-05

A review of abstract concept learning in embodied agents and robots.

Cangelosi, A

http://hdl.handle.net/10026.1/12412

10.1098/rstb.2017.0131
Philos Trans R Soc Lond B Biol Sci

PHILOSOPHICAL
TRANSACTIONS B

Phil. Trans. R. Soc. B.
doi:10.1098/not yet assigned

A Review of Abstract Concept Learning in Embodied Agents and Robots

Angelo Cangelosi[1*†] and Francesca Stramandinoli[2]

[1]*Centre for Robotics and Neural Systems, Plymouth University, Plymouth, PL4 8AA, UK, orcid.org/0000-0002-4709-2243*
[2]*iCub Facility Department, Istituto Italiano di Tecnologia, Via Morego 30, 16163, Genoa, Italy, orcid.org/0000-0002-6963-5931*

Keywords: Cognitive Robotics, Abstract Concept Learning, Numerical Concepts, Symbol Grounding, Embodied Cognition.

Summary

This paper reviews computational modelling approaches to the learning of abstract concepts and words in embodied agents such as humanoid robots. This will include a discussion of the learning of abstract words such as "use" and "make" in humanoid robot experiments, and the acquisition of numerical concepts via gesture and finger counting strategies. The current approaches share a strong emphasis on embodied cognition aspects for the grounding of abstract concepts, and a continuum, rather than dichotomy, view of concrete/abstract concepts differences.

Introduction

The robots' learning and understanding of abstract concepts and words, as well as concrete words related to the naming of objects and actions, can enable human users to shape the robots' behaviour by using one of the most natural interface at their disposal, i.e. language. In the future robots are expected to work as assistants for humans, performing joint tasks as co-workers in manufacturing scenarios, as housekeepers, and caregivers for elderly and people with disability. In such activities, is it important for the robot companions not only to be able to understand sentences such as "Lift the hammer", but also requests such as "Use the hammer" or "Use the plan". This paper proposes an approach to the learning of abstract concepts in robots by exploiting the grounding strategies currently used in the learning of the names of objects and actions, using embodied and situated strategies. Numerous models exist of how a robot can autonomously learn, i.e. ground, the association between concrete words and the corresponding objects seen and used by the robot, and the action it can perceive and act. This grounding can then further be transferred to higher-order, more abstract concepts such as "accept", "use", "make". Additional embodied strategies, such as gestures and finger use, can be exploited to acquire abstract numerical concepts.

Abstract words are used in daily conversations among people to communicate and describe events and situations that occur in their social and physical environment. Abstract words can be differentiated by concrete ones according to different criteria, ranging from concreteness, imageability and context availability, to mode of acquisition, etc. As claimed by Borghi and Binkofski (1) abstract words, with respect to concrete ones, are characterised by different grounding, complexity and meaning variability. A common way to distinguish between concrete and abstract concepts is to refer to their concreteness. Indeed, while concrete words refer to material and tangible entities that can be perceived through senses, abstract concepts have weaker perceptual constraints with sensorimotor experience and physical referents. However, the differentiation of words as concrete/abstract is a controversial problem. Evidence suggests a continuum, rather than dichotomy, view of concrete/abstract concepts (2). In addition, according to Altarriba et al. (3) words that refer to emotions should be categorised apart from concrete and abstract words (3). As suggested by Barsalou (4), concepts become increasingly abstract as they get more detached from physical entities, and more associated with mental states. For example, words that are purely definitional (e.g. "odd number") are

*Author for correspondence (a.cangelosi@plymouth.ac.uk).
†Present address: Centre for Robotics and Neural Systems, Plymouth University, Portland Square A316, Plymouth, PL4 8AA, Devon, UK

more abstract than words that refer to social roles (e.g. "student"), that in turn are more abstract than strongly concrete and perceivable entities (e.g. "book"). Further, words such "push" and "give" can be differentiated in their level of concreteness and motor modality; that is, a word like "push" is uniquely linked with the action of pushing by using the hand, while "give" implies multiple motor instances of the process of passing an object by using one hand, two hands, the mouth etc. Similarly, the meaning of words like "use" and "make" is general and depends on the context in which words occur (4). In a scenario in which a person is interacting with a set of tools, the meaning of "use" is specified by the particular tool employed during the interaction (e.g. "use a knife", "use a brush"), while the meaning of "make" depends on the outcome of interactions (e.g. "make a slice", "make a hole").

Abstract Concepts Representation and Grounding

Different views have been proposed on the representation of abstract concepts and words.
According to traditional views (e.g. (5)) both concrete and abstract concepts are represented as abstract and a-modal symbols that are unrelated to the perceptual states and actions that produce them. For example, according to the Context Availability Theory (6), the different processing involved for concrete and abstract concepts is due to the fact that concrete words have stronger semantic relations with a reduced number of contexts, while abstract words have weak relations with a larger number of contexts. Following the Dual Coding Theory (7), concrete concepts are represented by activating a verbal and non-verbal system, while abstract concepts are represented in a verbal system only.

The embodied view (e.g. (8)), by contrast, argues that concrete and abstract concepts are represented as modal symbols grounded in perception and action (i.e. sensorimotor knowledge). Within the embodied view, theories based on "simulations" (8), "metaphors" (9) and "actions" (10), (11) have been proposed.
A recent proposal (12) claims that conceptual processing requires the activation of multiple representational systems (i.e. sensorimotor and linguistic knowledge).

Given the current debate in the field and the complexity of the matter, the learning of abstract concepts and words is increasingly proving to be an extremely complex task for grounded theories of cognition as well as for embodied computational modelling and robotics. The traditional computational approach considers conceptual representations independent from perceptual and motor knowledge, neglecting the role of the experience made in the world through physical interaction. This approach has been criticized and challenged because of the lack of connection between symbols and their real-world referents, as with Searle's (13) famous *Chinese Room Argument*, and Harnad's (14) *Symbol Grounding Problem.* In the formulation of the symbol grounding problem, Harnad argued that purely computational symbols are self-referential entities that require the interpretation of an external experimenter to identify the referential meaning of lexical items.

In contrast to purely computational modelling methods, embodied approaches to language learning focus on the design of artificial cognitive agents that are capable to ground concepts and words by integrating perception and action and via direct experience and use of these words in a situated, embodied context. The robotics modelling of abstract words uses and extends methods from the current literature on the grounding of concrete words for objects and actions. Two main grounding mechanisms can be used: (i) "direct grounding", where the robot learns the names of objects it is perceiving, or words for actions it is performing or observing (15), (16); (ii) "grounding transfer" where new words are acquired via word combinations and without direct sensorimotor experience of their referents (17). For example, the word "unicorn" can be learned without ever seeing such an imaginary animal, but through the definition "unicorn is horse with horn", where the grounded meanings of "horse" and "horn", previously learned via direct grounding, can be transferred to the new word. Moreover, a third mechanism used to model abstract word learning in robots is that of combining gestures and action with words, such as in the use of finger counting to teach a child (or a robot) to count. Such motor strategies allow the learning agent to map abstract concepts as ordered numbers into embodied concepts as finger sequences.

Towards Abstract Words in Robots via Grounding Transfer

At the two extremes of the concrete/abstract continuum view of concepts, there are strongly concrete and perceivable words, such as "pencil" and "push", and very abstracts ones such as "democracy", "freedom". Between these two extremes it is possible to consider different levels of abstractness. For example, the word "accept" (as in "accept a present") is an extension of the concrete action of receiving an object but in a friendly social context. Whereas, the word "use" (as in "use a pencil") is a more abstract version of the concrete motor concept of "drawing with a pencil" (though still linked to action such as draw). In recent studies Stramandinoli et al., (18), (19), (20) have tackled the problem of grounding these intermediate abstract concepts (like "accept" and "use" defined as higher-order concepts), adopting the same grounding transfer mechanism, and implementing it in robot experiments. Their approach is based on the operationalization of

Barsalou's (8) theory of mental simulation and conceptual combination for the acquisition of higher-order concepts. This simulation theory is implemented via the symbol grounding transfer method proposed by Cangelosi and Riga (17), which requires the implementation of the (i) basic grounding (BG) and (ii) higher grounding transfer (HG) mechanisms. During the basic grounding, the robot learns to ground a set of action primitives (e.g. "push", "pull", "grasp"); whereas, during the grounding transfer linguistic descriptions consisting of a sequence of words (e.g. "receive is push, grasp and pull" provided in the form a simplified linguist token) guide the hierarchical organization of the basic concepts directly grounded in sensorimotor experience (e.g. "push", "pull", "grasp") in order to learn novel concepts (e.g. "receive"). The grounding transfer mechanism enables the robot to create the semantic reference for higher-order words that do not have a direct and tangible relation with sensorimotor experience.

The concrete/abstract continuum view of concepts presented above, and Barsalou's simulation theory, are examined through developmental robotics models of the direct grounding and grounding transfer mechanisms on sensorimotor knowledge (17). Stramandinoli et al., (19) have performed experiments on a simulated environment for the iCub robot (21), (22), adopted as a study platform for research studies on the grounding of abstract words in cognitive robots. This developmental modelling approach, which will be used in the computational models reviewed below, follow the principles of Developmental Robotics (23). This aims at the building of cognitive robot models and experiments which take direct inspiration from developmental psychology phenomena on sensorimotor and cognitive development, such as stage-like patterns of developmental changes and open-ended, cumulative learning.

The robot's cognitive model is based on Recurrent Neural Networks (RNNs) that permit the learning of higher-order concepts based on temporal sequences of action primitives and word sentences. RNNs, such as the Elman simple recurrent network, are particularly suitable for modelling abstract concept learning since the recurrent connections allow the network to handle time series and sequences of times/words. This is the case, for example, in using sequences the combination of words when defining a composite, abstract concept, and in the case of counting behavior for learning abstract number concepts.

The training of the model is incremental. During the BG training the robot learns the names associated to the action primitives through direct sensorimotor experience. The names of action primitives, given in input to the robot's neural network, are "push", "pull", "grasp", "release", "smile", "frown", and "neutral". The words "smile", "frown", and "neutral" are not used to describe emotional states of the robot but rather motor acts. For the HG training, to enable different levels of combination of basic and complex actions, two different stages are implemented. In the first HG stage (i.e. HG-1), the robot learns three new higher-order words (i.e. "give", "receive", "pick") by combining only basic action primitives (i.e. "receive is push, grasp and pull"). In order to obtain the transfer of grounding from basic actions to higher-order words, the network calculates separately the output corresponding to the words contained in the description ("push", "grasp", "pull") and stores it. Subsequently, the network receives as input the higher-order word "receive" and as target the outputs previously stored. During the second HG stage (i.e. HG-2), the robot learns three new higher-order words ("accept", "reject", "keep") consisting of the combination of basic action primitives and higher-order words acquired during the previous HG-1 stage (e.g. "accept is receive and smile"). HG-2 adds a further hierarchical combination of words from both concrete concepts (BG) and first level of abstraction words (HG-1). This training methodology is extremely flexible and permits to freely add novel words to the known vocabulary of the robot, or to completely rearrange the word-meaning associations.

Similarly, in Stramandinoli et al., (20) experiments on the iCub robot where performed for investigating the grounding of abstract action words. Indeed, the grounding transfer mechanism is used for the learning of words with more general, abstract meanings, such as "use" and "make", referred as abstract action words. Abstract action words represent a class of terms distant from immediate perception that describe actions with a general meaning and that can be referred to several events and situations. Therefore, they cannot be directly linked to sensorimotor experience through a one-to-one mapping with their physical referents in the world. The grounding of abstract action words is achieved through the integration of the linguistic, perceptual and motor input modalities, recorded from the iCub sensors, in a three layers RNN model. The iCub robot first develops some basic perceptual and motor skills, necessary for initiating the interaction with the environment, and then it can use such knowledge to ground language. The training of the model consists of the following incremental stages: (i) pre-linguistic, (ii) linguistic-perceptual and (iii) linguistic abstract.
During the pre-linguistic training, the iCub acquires a set of basic visual and motor skills leveraged for: (i) extracting objects features (i.e. dimension, colour and shape) and (ii) performing basic motor primitives (i.e. "push", "pull", "lift", "lower", etc.). Object features are extracted from the visual stream read from the iCub cameras while the robot interacts with the toy objects such as "knife", "saw", "pencil" "brush". Hence, for each object a 4x4 binary matrix that represent the extracted features is created. Furthermore, the robot is endowed with a set of basic motor primitives such as "push", "pull", and "lift" that enables it to initiate its physical interaction with the environment. Through the combination of motor primitives into sequences, the robot can learn to perform action primitives that correspond to more complex behaviors. Indeed, action

primitives (e.g. "cut", "paint") are built by combining low level motor primitives together. For example, the action primitive "cut" is built by iterating the "push-pull" sequence several times.
The linguistic-perceptual training is the first stage of language acquisition. The robot is trained to name actions performed with objects (two-words sentences consisting of a verb followed by a noun e.g. "cut with a knife"); these words are directly grounded in perception and motor experience. The model, which was previously trained to extract object's features and perform action primitives, during this stage associates labels to the corresponding object and actions.
During the linguistic-abstract training, abstract action words (i.e. "use" and "make") are grounded by combining and recalling the perceptual and motor knowledge previously linked to basic words (i.e. linguistic-perceptual training). To derive the meaning of abstract action words the robot, guided by linguistic instructions (e.g. "use a knife"), organizes the knowledge directly grounded in perception and motor knowledge. This phase of the training represents the abstract stage of language acquisition when new concepts are formed by combining the meaning of terms acquired during the previous stages of the training. Novel lexical terms can be continually acquired throughout the course of the robot's development through new sensorimotor interactions with the environment to which correspond new linguistic descriptions.
At the end of the training, the robot is able to perform the behavior triggered by the linguistic description and the perceived object.

Learning Numerical Concepts in Robots via Gestures

Number cognition, such as the understanding and use of number words (e.g. one, two, twenty) and of fuzzy quantifies (few, some, many), and the manipulations of these numbers (from addition and multiplication to complex calculus) are another key example of abstract concepts and how embodiment strategies are used in the early developmental stages. Various embodied strategies, such as pointing and counting gestures, object touching, finger counting, and mathematical educational strategies based on spatial metaphors, have been shown to facilitate the development of number cognition skills (e.g. (24), (25)). The embodied basis of numbers is also evident in adults, such as with the size, distance and SNARC effects (Spatial-Number Association of Response Codes; (26)). This link between embodiment and early number learning has been exploited to teach robot numerical words and concepts, for two specific examples of pointing gesture whilst counting and on finger counting.

The role of pointing and touching gestures in the acquisition of counting skills is a prototypical developmental phenomenon from the point of view of the embodiment of linguistic and symbolic knowledge. When learning to count, children spontaneously point to, touch, or move objects, and a wide set of studies exists which demonstrate the beneficial effect of sensorimotor strategies on counting performance (see (27) and (24) for reviews). There are three main groups of hypotheses on the role and mechanism behind this phenomenon. First, gestures may help the child overcome the limitations in limited cognitive resources, for instance by helping her to keep track of counted items. Second, gestures may perform a coordinative function by combining a temporal correspondence with speech and a spatial correspondence with the counted items. Third, gestures may also facilitate social learning by providing the tutor with feedback on the child's learning progress.

Rucinski et al. (28) has proposed a developmental robotics model of the contribution of the counting gestures to learning to count. The robot experiments were modelled on Alibali & DiRusso's (23) study of the role of counting gestures in children, in particular the condition when the child sees a puppet pointing at the objects being counted aloud. The robot was trained and tested in several experimental conditions, varying the availability of the sensory signals (vision and gestures) to the robot, and the type of the counting gestures. The robot's control architecture consists of an Elman simple recurrent neural network. Thus, the counting task was simulated as requiring the network to output a count list corresponding to the number of objects shown in the visual input layer in response to the counting trigger stimuli (with the option of seeing natural counting gestures, where the robot sees the tutor's virtual hand pointing at each object). To investigate the importance of the spatial correspondence which characterises the "natural" counting gestures, Rucinski et al. (27) further contrasted such gestures with "artificial counting gestures". These consists of the rhythmic swings of the virtual arm back and forth, in which the gesture still corresponds to the recited number words in the temporal domain, but, unlike in "natural" counting gestures, it does not correspond to the counted items in the spatial domain.

The results of the robot experiments consistently show that the perception of the pointing gestures allows the robot to significantly improve the counting accuracy, as compared with the condition of counting using only visual information. This improvement was not explained simply by the additional input signal, as the model also counted significantly worse if it was given only he proprioceptive input. This provided first evidence outside of behavioural studies that counting gestures are a useful embodied cue in learning to count. Furthermore, contrasting the effects of natural spatio-temporal counting gestures with those of artificial rhythmic ones revealed that it is essential that the counting gestures are characterized by a spatial correspondence to the counted items – in the latter case the gestures did not facilitate the extraction of information by the neural network from the visual input. Whereas "natural" counting gestures enabled the neural network to extract more information from the visual input, this was not the case with the "rhythmic" gestures. In the rhythmic condition, the robot's neural network converged into counting the gestures rather than the objects, as the counting performance was indistinguishable with, and without, the visual information for this type of gestures.

Such a pioneering model of the learning of abstract numerical concepts in robots by exploiting the role of pointing gesture has been more recently complemented by a study on finger counting, another key embodiment strategy extensively used by children to learn to count. Several psychology and neuroscience studies with children and adults show that finger counting strategies and finger-based representations play an important role in the development of numerical and arithmetical skills and in the learning of number words. Moreover, finger counting in particular, and gesture and action-based embodied strategies in general, have been shown to support more effective acquisition of number words (e.g. (24)) and to affect the teaching of mathematical concepts (e.g. (25)).

The developmental robotics paradigm was used specifically to explore whether finger counting and the association of number words to each finger could serve to bootstrap the representation of number in the humanoid robot iCub (29), (30). This model uses a combination of neural networks, with a recurrent intermediate layer for number words and motor finger sequences, to implement the learning of associations between (motor) finger counting, (visual) object counting and (auditory) number words and sequence learning. The study manipulates the coupling between different modalities. In the Auditory-Only condition, the robot only learns to hear and repeat the sequence of number words ("one", "two", … up to "ten"). In the Finger-Only condition, the iCub is trained to produce finger counting sequences, without any auditory signal in input or output. Finally. in the Finger+Auditory condition, the robot simultaneously learning the sequence of acoustic number words and the sequence of moving fingers. The American sign language finger counting configuration was used to match the iCub robot's finger actuator system.

The results obtained in various experiments with both the simulated and the physical iCub robot show that learning the number word sequences together with finger sequencing helps the fast building of the initial representation of number in the robot. Robots who only learn the auditory sequences perform worse. Moreover, the neural network's internal representations for these two counting conditions result in qualitatively different patterns of the similarity between numbers. Only after the Finger+Auditory sequence learning does the network represent the relative distance between them, which corresponds to the quantitative difference between numbers. In Finger+Auditory trained robots, the cluster analysis diagram of the hidden layer's activation shows that the representation for the number word "one" is adjacent to that of "two" and is increasingly different (distant) from the higher numbers. Instead, in the auditory-only condition, there is no correspondence between the cluster diagram similarity distance and the numerical distance. Moreover, this modelling approach has been extended to simulate (30) the acquisition of different cultural strategies in finger counting.

Conclusions

This paper reviews the very first, pioneering robotic models of abstract concepts learning, all adopting an embodied methodological approach towards the grounding of abstract words and numbers and sharing some common methodological and theoretical principles.

One common characteristic of these approaches is that the robot's cognitive architecture is based on recurrent neural networks to process and ground abstract meanings, allowing the implementation of embodied cognition theories on mental simulations and symbol grounding. Recurrent architectures permit both the simultaneous, multimodal processing of sequences or words and actions (pointing gestures, finger counting), and the processing of linguistic sentences for the composition of their meanings. In the specific case of the abstract word experiments, the implementation of the grounding transfer via recurrent networks has been interpreted as an operationalization of Barsalou's (8) mental simulation and conceptual combination mechanisms (17). Moreover, the direct grounding of highly concrete motor primitive words ("push", "pull"), combined with the learning of new, higher order action words (e.g. "accept", "use", which have an increasingly more generic – i.e. towards abstractness - sensorimotor meaning), achieved via linguistic definitions, is in line with theories on the activation of multiple representational systems via sensorimotor and linguistic knowledge (12). Further, the use of neural-network based architectures also has the advantage of modelling the representation of concrete and abstract meanings not via arbitrary, researcher-defined symbolic representations, but rather as learned parallel distributed representations combining perceptual, motor and linguistic knowledge. And as the link between linguistic and embodied meaning is autonomously learned by the neural network, such an approach and architecture satisfies the symbol grounding problem.

The approach reviewed here is further consistent with other embodied theories of cognition. For example, in the finger counting experiments the robot's reliance on the constraints of the hand structure and finger counting sequence to develop representation of number words consistent with number cardinality, can provide an operational implementation of the mechanisms of metaphors between body/space maps and numbers (9).

A further common feature of the models and experiments reviewed above is the use of a developmental robotics approach (23). This implies the modelling of various developmental stages and capabilities involved in the grounding of concrete and abstract words and numbers, taking inspiration from child psychology. The finger counting studies clearly relies on a common, universal strategy on number learning associated to finger counting. The Rucinski number (28) and pointing gesture model directly simulates Alibali and DiRusso (24) experiments on the children's use of pointing and touching whilst counting.

The field of the modelling of abstract concept learning in robots and machines, using embodied strategies and representations, still is at its "infancy", to use a developmental metaphor. Notwithstanding the significant achievements of the above studies to show the possible acquisition of a small set of concrete and abstract words in robotic agents, big challenges lie ahead both in terms of the complexity of the lexicon, and the level of abstractness of the words used. However, within the view a continuum between concrete and abstract concepts, such small steps and models provide methodological insights and computational solutions for the big challenge of designing robots capable to interact with humans and hold meaningful conversations on the abstract concept of "using a plan" and talk about "freedom".

Funding
This work was supported by the Marie Curie Intra European Fellowship RoboTAsk (624424) within the 7th European Community Framework Programme (FP7), and by the Marie Curie ITNs RobotDoc (624424) and H2020 APRIL (674868).

References

1. **Borghi, Anna M. and Binkofski, Ferdinand.** *Words As social Tools: An embodied view on abstract concepts (Briefs in Cognition series).* New York : Springer, 2014.
2. *Imagery, context availability, contextual constraint, and abstractness.* **Wiemer-Hastings, Katja, Krug, Jan and Xu, Xu.** 2001. Proceedings of the 23rd annual conference of the cognitive science society. pp. 1134-1139.
3. *Concreteness, context availability, and imageability ratings and word associations for abstract, concrete, and emotion words.* **Altarriba, Jeanette, Bauer, Lisa M. and Benvenuto, Claudia.** s.l. : Springer, 1999, Behavior Research Methods, Instruments, & Computers, Vol. 31, pp. 578-602.
4. *Abstraction in perceptual symbol systems.* **Barsalou, Lawrence W.** s.l. : The Royal Society, 2003, Philosophical Transactions of the Royal Society of London B: Biological Sciences, Vol. 358, pp. 1177-1187.
5. **Fodor, Jerry A.** *Concepts: Where cognitive science went wrong.* s.l. : Oxford University Press, 1998.
6. **Schwanenflugel, P. J.** Why are abstract concepts hard to understand? [book auth.] Schwanenflugel P.J. (Ed.). *The psychology of word meanings.* Hillsdale : Lawrence Erlbaum Associates, Inc, 1991, pp. 223-250.
7. **Paivio, Allan.** *Mental representations: A dual coding approach.* s.l. : Oxford University Press, 1990.
8. *Perceptions of perceptual symbols.* **Barsalou, Lawrence W.** s.l. : Cambridge Univ Press, 1999, Behavioral and brain sciences, Vol. 22, pp. 637-660.
9. **Lakoff, George and Johnson, Mark.** *Metaphors we live by.* s.l. : University of Chicago press, 2008.
10. *Grounding language in action.* **Glenberg, Arthur M. and Kaschak, Michael P.** s.l. : Springer, 2002, Psychonomic bulletin & review, Vol. 9, pp. 558-565.
11. *Processing abstract language modulates motor system activity.* **Glenberg, Arthur M., et al.** s.l. : Taylor & Francis, 2008, The Quarterly Journal of Experimental Psychology, Vol. 61, pp. 905-919.
12. *The linguistic and embodied nature of*

conceptual processing. **Louwerse, Max M. and Jeuniaux, Patrick.** s.l. : Elsevier, 2010, Cognition, Vol. 114, pp. 96-104.
13. *Minds, brains, and programs.* **Searle, John R.** s.l. : Cambridge Univ Press, 1980, Behavioral and brain sciences, Vol. 3, pp. 417-424.
14. *The symbol grounding problem.* **Harnad, Stevan.** s.l. : Elsevier, 1990, Physica D: Nonlinear Phenomena, Vol. 42, pp. 335-346.
15. *Integration of action and language knowledge: A roadmap for developmental robotics.* **Cangelosi, A., et al.** 3, 2010, Autonomous Mental Development, IEEE Transactions on,, Vol. 2, pp. 167-195.
16. *Grounding words in perception and action: computational insights.* **Roy, Deb.** s.l. : Elsevier, 2005, Trends in cognitive sciences, Vol. 9, pp. 389-396.
17. *An embodied model for sensorimotor grounding and grounding transfer: Experiments with epigenetic robots.* **Cangelosi, Angelo and Riga, Thomas.** s.l. : Wiley Online Library, 2006, Cognitive science, Vol. 30, pp. 673-689.
18. *Towards the grounding of abstract words: a neural network model for cognitive robots.* **Stramandinoli, Francesca, Cangelosi, Angelo and Marocco, Davide.** 2011. Neural Networks (IJCNN), The 2011 International Joint Conference on. pp. 467-474.
19. *The grounding of higher order concepts in action and language: a cognitive robotics model.* **Stramandinoli, Francesca, Marocco, Davide and Cangelosi, Angelo.** s.l. : Elsevier, 2012, Neural Networks, Vol. 32, pp. 165-173.
20. *Making sense of words: a robotic model for language abstraction.* **Stramandinoli, Francesca, Marocco, Davide and Cangelosi, Angelo.** s.l. : Springer, 2016, Autonomous Robots, Vol. 41, pp. 367–383.
21. *The iCub humanoid robot: an open platform for research in embodied cognition.* **Metta, Giorgio, et al.** 2008. Proceedings of the 8th workshop on performance metrics for intelligent systems. pp. 50-56.
22. *An open-source simulator for cognitive robotics research: the prototype of the iCub humanoid robot simulator.* **Tikhanoff, Vadim, et al.** 2008. Proceedings of the 8th workshop on performance metrics for intelligent systems. pp. 57-61.
23. **Cangelosi, Angelo, Schlesinger, Matthew and Smith, Linda B.** *Developmental robotics: From babies to robots.* s.l. : MIT Press, 2015.
24. *The function of gesture in learning to count: More than keeping track.* **Alibali, M. W. and DiRusso, A. A.** 1, 1999, Cognitive Development, Vol. 14, pp. 37–56.
25. *Effects of finger counting on numerical development – the opposing views of neurocognition and mathematics education,.* **Moeller, K., et al.** 328, 2011, Frontiers in Psychology, Vol. 2. doi: 10.3389/ fpsyg.2011.00328.
26. *The mental representation of parity and number magnitude.* **Dehaene, S., Bossini, S. and Giraux, P.** 1993, Journal of Experimental Psychology: General, Vol. 122, pp. 371–396.
27. *The role of gesture in children's learning to count.* **Graham, T. A.** 4, 1999, Journal of Experimental Child Psychology, Vol. 74, pp. 333–355.
28. *Robotic model of the contribution of gesture to learning to count.* **Rucinski, M., Cangelosi, A. and Belpaeme, T.** 2012. In Proceedings of the IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-Epirob 2012). pp. 1-6.
29. *Making fingers and words count in a cognitive robot.* **De La Cruz, V.M., et al.** 13, 2014, Frontiers in Behavioral Neuroscience, Vol. 8.
30. *Grounding Fingers, Words and Numbers in a Cognitive Developmental Robot.* **Di Nuovo, A., De La Cruz, V.M. and Cangelosi, A.** 2014. In: IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB). pp. 1-6.
31. *Learning adjectives and nouns from affordances on the icub humanoid robot.* **Yürüten, Onur, et al.** 2012. International Conference on Simulation of Adaptive Behavior. pp. 330-340.
32. *Learning semantic combinatoriality from the interaction between linguistic and behavioral processes.* **Sugita, Yuuya and Tani, Jun.** s.l. : Sage Publications Sage CA: Thousand Oaks, CA, 2005, Adaptive Behavior, Vol. 13, pp. 33-52.
33. *The emergence and evolution of linguistic*

structure: from lexical to grammatical communication systems. **Steels, Luc.** s.l. : Taylor & Francis, 2005, Connection science, Vol. 17, pp. 213-230.
34. *An abstract to concrete shift in the development of biological thought: The insides story.* **Simons, Daniel J. and Keil, Frank C.** s.l. : Elsevier, 1995, Cognition, Vol. 56, pp. 129-163.
35. *Semiotic schemas: A framework for grounding language in action and perception.* **Roy, Deb.** s.l. : Elsevier, 2005, Artificial Intelligence, Vol. 167, pp. 170-205.
36. *Discovering communication.* **Oudeyer, Pierre-Yves and Kaplan, Frédéric.** s.l. : Taylor & Francis, 2006, Connection Science, Vol. 18, pp. 189-206.
37. *Real-time parallel processing of grammatical structure in the fronto-striatal system: A recurrent network simulation study using reservoir computing.* **Hinaut, Xavier and Dominey, Peter Ford.** s.l. : Public Library of Science, 2013, PloS one, Vol. 8, p. e52946.
38. *Real-time spoken-language programming for cooperative interaction with a humanoid apprentice.* **Dominey, Peter Ford, Mallet, Anthony and Yoshida, Eiichi.** s.l. : World Scientific, 2009, International Journal of Humanoid Robotics, Vol. 6, pp. 147-171.
39. *Simulating the evolution of language.* **Cangelosi, Angelo and Parisi, Domenico.** s.l. : Springer-Verlag New York, Inc, 2002.
40. *Short-term memory for serial order: a recurrent neural network model.* **Botvinick, Matthew M. and Plaut, David C.** s.l. : American Psychological Association, 2006, Psychological review, Vol. 113, p. 201.
41. *Grounding speech utterances in robotics affordances: An embodied statistical language model.* **Stramandinoli, Francesca, et al.** 2016. Development and Learning and Epigenetic Robotics (ICDL-EpiRob), 2016 Joint IEEE International Conference on. pp. 79-86. (http://dx.doi.org/10.1109/DEVLRN.2016.7846794).
42. **Gibson, James J.** *The ecological approach to visual perception: classic edition.* s.l. : Psychology Press.
43. *A Deep Learning Neural Network for Number Cognition: A bi-cultural study with the iCub.* **Di Nuovo, A., De La Cruz, V.M. and Cangelosi, A.,.** s.l. : IEEE, 2015. In Joint IEEE International Conference on Development and Learning and Epigenetic Robotics (ICDL-EpiRob), 2015 . pp. 320-325.